

Memory and Subjective Workload Assessment

Lowell Staveland - San Jose State University Foundation, San Jose, CA
Sandra Hart - NASA-Ames Research Center, Moffett Field, CA
Yei-Yu Yeh - University of Illinois, Urbana-Champaign, ILL

Abstract

Recent research suggested subjective introspection of workload is not based upon specific retrieval of information from long-term memory, and only reflects the average workload that is imposed upon the human operator by a particular task. These findings are based upon global ratings of workload for the overall task, suggesting that subjective ratings are limited in ability to retrieve specific details of a task from long-term memory. To clarify the limits memory imposes on subjective workload assessment, the difficulty of task segments was varied and the workload of specified segments was retrospectively rated. The ratings were retrospectively collected on the manipulations of three levels of segment difficulty. Subjects were assigned to one of two memory groups. In the Before group, subjects knew before performing a block of trials which segment to rate. In the After group, subjects did not know which segment to rate until after performing the block of trials. The subjective ratings, RTs, and MTs were compared for within group, and between group differences. Performance measures and subjective evaluations of workload reflected the experimental manipulations. Subjects were sensitive to different difficulty levels, and recalled the average workload of task components. Cueing did not appear to help recall, and memory group differences possibly reflected variations in the groups of subjects, or an additional memory task.

Introduction

Much attention is being focused on the utility of subjective evaluations to measure mental workload and human performance. The potential for subjective ratings to reflect a human operators sensitivity to varying task demands, has been validated in several experiments (Yeh, Wickens & Hart 1985; Hart, Sellers, & Guthart, 1984; Arbak, Shew, & Simons 1984). These findings, however, are based on global ratings of workload for a group of similar tasks, or segments of a continuously changing task (Bortolussi, Kantowitz, Hart, 1985), which measure the overall loading on cognitive processes, irregardless of when they were obtained. Global ratings obtained while performing a task are highly correlated with the global ratings obtained retrospectively (Bortolussi et al, 1985), even though they may not reflect moment-to-moment variations in cognitive loads that operators experience while performing a task. Yeh et al, (1984) found that "...subjective introspection of workload is not based on specific retrieval of information from working memory and only reflects the average workload imposed on human operators by a particular task".

The tasks selected for their study were based on the 'Fittsberg' paradigm (Hartzell et al) which was originally based on the serial combination of FITTS target aquisition tasks following selection among the alternative locations based on a STERNberg memory search decision. For this application, two response selection tasks were used: pattern match and arithmetic equations. For each response selection task and target aquisition task, three levels of difficulty were imposed. Difficulty levels of the two task components were

consistent within a block of trials, and either both were increased or decreased in difficulty, or the difficulty of one component was increased while the other decreased. Measures of performance independently reflected task difficulty manipulations within trial blocks; RT varied with RS difficulty, whereas MT varied with RE difficulty. Workload ratings accurately reflected the integrated workload of all tasks within a block, displaying no primacy/recency effect, or greater influence by one task component than another. Since ratings were consistently equal to the average workload of a block of trials, the question remained whether subjects were simply insensitive to task manipulations, or in fact accomplished the summary evaluation that was required by the design of the experiment. In either case, it was not clear whether subjects would have been able to provide more selective evaluations of trial block segments had they been required to do so. Such global ratings are fine where the goal is to evaluate differences between tasks (e.g. comparing the difficulty of one flight to another). In many circumstances though, the difficulty of specific segments within a flight need to be evaluated. In this case global ratings do not suffice. More detailed evaluations are required to reflect the varying difficulty levels experienced by operators during a flight.

Previous research suggested that delaying retrospective evaluations of task segments does not significantly alter the relationships among reflective ratings, even though the absolute values might be somewhat different (Eggemeier, Melville, & Crabtree 1984; Notestine 1984). Even intervening task performance does not significantly effect workload ratings (Eggemeier, et al 1984). These results have direct implications for this study, considering subjects had to reflectively rate different segments of a task after a block of segments. If a subject is asked to rate the first segment out of three in a block of trials, the intervening segments should not significantly effect their retrospective rating. This means the workload ratings obtained in this study should reflect specific retrieval of a particular segment from long-term memory, independent of the other segments influence on ratings. Delays in rating the first or second segments while performing the second or third segments also should not influence subjective experience of workload. This rules out delay as a confounding variable, and increases the confidence in the obtained ratings as being indicative of an operators workload and cognitive loading for a particular segment.

The current study addressed the limits memory imposed on subjective ratings. Subjects were divided into two memory groups: Before and After. Subjects in the Before group knew in advance the segment-to-be-rated. Subjects in the After group did not know in advance the segment-to-be-rated, they were told after completing the block of trials which segment to rate. The purpose was to elicit answers to the following questions: (1) How sensitive are subjects to task component manipulations? (2) Is the information about different segments in a task available retrospectively? Or is the average workload all that can be recalled (3) Does knowing in advance the segment-to-be-rated aid recall? And (4) Do all task components contribute equally to workload? This experiment follows up Yehs findings that subjective ratings are limited in their capacity to retrieve specific details from working memory.

The task selected for this experiment was based on a version of the Fittsberg paradigm used by Yeh et al (1985), and Hartzell et al, (1983). It involved two components: response selection and response execution. The response selection component was based on completing arithmetic equations. As

the equations complexity increased from one operator to three, difficulty increased as well. The response execution component was a target acquisition task based on Fitts law (Fitts & Petersen, 1964). Its difficulty was manipulated by varying the targets index of difficulty (ID). The two components were combined to form three categories: Consistent: The RS/RE components had a consistent difficulty level across the three segments within a condition; (2) Changing-consistent: RS/RE components difficulty levels were positively correlated, either increasing or decreasing in difficulty from segment to segment within a condition; and (3) Changing-inconsistent: RS/RE components difficulty levels were negatively correlated (the RS component increased while the RE component decreased, or vice-versa). Cognitive loading was expected to vary as a function of the response selection component, whereas response execution would influence MTs. Workload ratings were expected to vary as a joint function of the difficulty levels of both components within each trial-block segment.

Method

Subjects

Eighteen male and two female subjects served as paid volunteers. None had any prior experience with Fitts tasks, but all had served as subjects in other experiments at NASA-Ames Research Center. Thus, most had experience with the use of the bipolar rating scales. All subjects had competent arithmetic skills.

Apparatus

The experiment was conducted in a sound-attenuated chamber. The subject was seated in a chair located 85 cm from a 23-cm monitor where all experimental tasks were displayed. The visual angle subtended by the most extreme targets was 11 deg. A two-axis joystick was mounted on the right arm of the chair for response selection and target acquisition responses. Subjective ratings were entered with a slide pot and button mounted on the left arm of the chair. The experiment, data acquisition, and reduction were performed with an Apple II+ microcomputer, modified to allow rapid recording of response (10 msec resolution). The data were analyzed with a Dec 11/70, and a Vax 11/750.

Task Components

Each task had two components: response selection and response execution. The outcome of the response selection task served as input to the response execution task. Thus, the two task components could be performed serially and were functionally related. There were three levels of difficulty for each component: easy (E), medium (M), and hard (H). The two components were combined to form seven conditions: EE, MM, HH, II, DD, ID, DI. The first letter of each pair represents the response selection component, and the second letter for the response execution component. 'I' indicates that the difficulty of that component was increased from the beginning to the end of that trial block; 'D' indicates that it decreased.

Response Selection The solution to an equation performed mentally

determined the direction of movement. Each equation involved one, two or three mathematical operations which determined the level of difficulty. The easy condition required one operation, (e.g. 2+3.), medium required two (e.g. 3*2/1), and hard required three (e.g. (4-1)*3). The solutions were always whole numbers, either greater or less than a single digit memory set presented prior to each block of trials. These were similar to three of the RS tasks employed in the previous study (Yeh et al, 1985). Subjects were told to move the joystick right if the solution was greater than the remembered digit (7, 8, or 9), or left if it was less. The interval between stimulus onset and a 2% joystick deflection was recorded as reaction time (RT).

Response execution. The response execution component was a target acquisition task. Two identical target areas were displayed symmetrically on either side of the stimulus at a distance determined by the index of difficulty computed according to Fitts law ($ID = \log_2(2A/W)$). The targets were two 1.25 cm lines separated by a distance appropriate for the ID of that condition. The same ID levels used in earlier studies were selected for the three levels of difficulty: Easy = 2.52, Medium = 4.19, and Hard = 5.67. The interval between a 2% joystick deflection and satisfaction of the steadiness criterion for keeping the cursor within the target, was recorded as movement time (MT).

Condition Characteristics

Each of the seven experimental blocks of trials (EE, MM, HH, II, DD, ID, DI) were divided into three equal segments of twelve trials each. The eight equations within a segment had the same difficulty level as the eight IDs, but the difficulty levels from one segment to the next depended on the condition. For EE, MM, and HH conditions, all three segments within a block had the same response selection and target acquisition difficulty levels (consistent). For two other conditions (changing-consistent), the difficulty of both components either increased (II) or decreased (DD). For the last two conditions, (changing-inconsistent), the difficulty of the two components, (ID, and DI), changed in opposite directions. The six equations that transitioned between segments were randomly mixed so that the divisions between segments was less evident. Capture time (RT+MT), was the total response time for each trial, averaged across all trials, and was presented as feedback at the end of each condition along with the number of correct responses.

Subjective Ratings

Two types of ratings were collected in this study:

(1) Individual differences in definition. The relative importance of nine factors to each subject's definition of mental workload was determined. These nine factors were: task difficulty, time pressure, own performance, physical effort, mental effort, frustration, stress, fatigue, and activity type (Yeh et al, 1985). Each factor was paired with every other factor (36 pairs) in a pretest. Subjects, selected the member of each pair that was most related to their definition of workload. Each factor could be selected from 0 (never considered relevant) to 8 (more important than any other factor) times. The number of times a factor was selected was its weight.

(2) Bipolar ratings. Ratings on nine bipolar rating scales plus an overall workload scale were collected at the end of each condition. Each scale was presented on the experimental display as an 11 cm vertical line with a title (e.g., "OVERALL WORKLOAD) and bipolar descriptions at each end (e.g., "EXTREMELY HIGH/EXTREMELY LOW"). The cursor was positioned at the desired point on the scale with a slide pot, and entered with a button. Each selection was assigned a value from 1 to 100 during data reduction.

Procedure

Each subject participated in the experiment two hrs per day, for three days. The first day, and the first 30 min on subsequent days were used for practice.

The subjects read a brief explanation of the experiment to familiarize themselves with the objectives and experimental tasks. After the workload weights were collected, the subjects practiced the target acquisition task: 20 blocks of 24 trials each. The basic response execution task entailed acquiring a target displayed on either the right or left side of the display; there was no response selection task. Following this, they performed the three difficulty levels of the response execution task (E,M,H), the response selection task (E,M,H): no targets were displayed, and the combined tasks (E,M,H). The response selection task entailed solving an equation, and moving the joystick right if the solution was greater than the remembered digit, or left if the solution was less. The practice trials at the beginning of each subsequent day were combined tasks involving changing-consistent (II,DD), and changing-inconsistent (ID,DI) conditions.

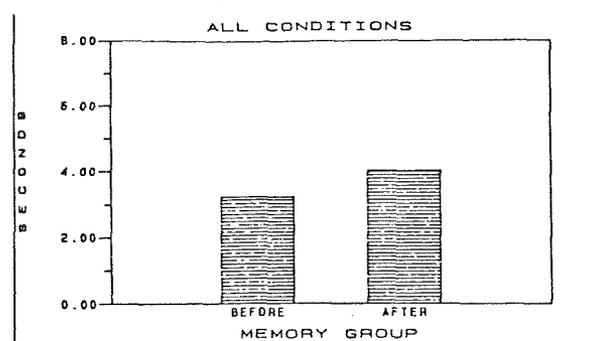
Each of the seven conditions were presented three times, so subjects could rate the workload of the first twelve trials after one block, the second twelve trials after another, and the third twelve after the third block. Subjects in the before group were told the segment-to-be-rated before performing each block of 36 trials. Subjects in the after group were told the segment-to-be-rated after performing each block of 36 trials. A total of 21 experimental conditions were rated. The segments-to-be-rated were presented to each subject in counterbalanced order, and the seven different conditions were presented in random order.

Results

General Comparison of Memory Groups

ANOVAs of mean RTs and MTs, percent correct, and bipolar ratings were collected for each of the three segments for the seven conditions in the three categories: consistent, changing-consistent, and changing-inconsistent. As shown in Figure 1a, the RTs for the Before group were less than for the RTs of the After group. RTs reflected the response selection difficulty, and were not affected by response execution difficulty. MTs for the Before group

Figure 1a. RT-Before vs After for all conditions.



were greater than the After group, and reflected response execution difficulty, but did not reflect response selection difficulty (Figure 1b). The MT, and RT results were consistent across all conditions for both experiments. RTs were always greater than MTs. The average levels of workload ratings were similar for the two groups. However, differences in response to experimental manipulations were observed.

Percent Correct

There were no significant speed-accuracy trade-offs. In the consistent condition, there was a trend for both speed and accuracy to decrease, as the difficulty increased from conditions 'EE' to 'MM' to 'HH'. For the changing-consistent, and changing-inconsistent conditions, this trend is not apparent between conditions, or between segments. Overall, the subjects were highly accurate across all conditions and segments, $F(1,9) = 534.03, p < .001$.

RTs and MTs.

The ANOVA results for the Before and After groups are presented in Figures 2a-2c, 3a-3c, and 4a-4c.

Consistent. RTs and MTs reflected the relevant RS or RE difficulty manipulations, (Figure 2a). The Before RTs were less than the After ($F(1,486) = 27.95, p < .001$) (Figure 2b). The Before MTs were greater than After ($F(1,486) = 35.52, p < .001$), (Figure 2c).

Before group. RT increased as the math equations increased in complexity (EE to MM to HH) ($F(2,18) = 32.1, p < .001$), reflecting an increase in cognitive loading. MTs also reflected these results, increasing in duration as RE difficulty increased from (EE to MM to HH) ($F(2,18) = 68.51, p < .001$).

After group. The results followed the same pattern as the Before group. RTs increased as RS difficulty increased across the three conditions (EE,MM,HH) ($F(2,18) = 87.88, p < .001$).

Figure 1b. MT-Before vs After for all conditions.

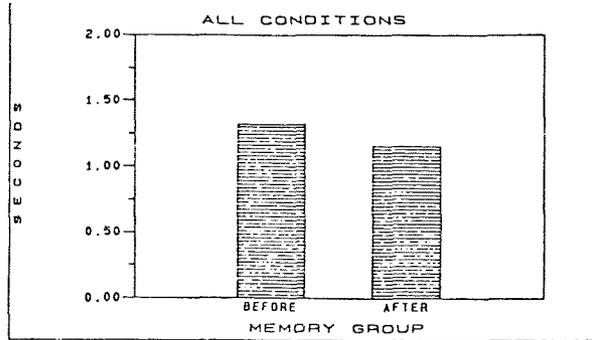


Figure 2a. Capture time-RT vs MT for consistent conditions.

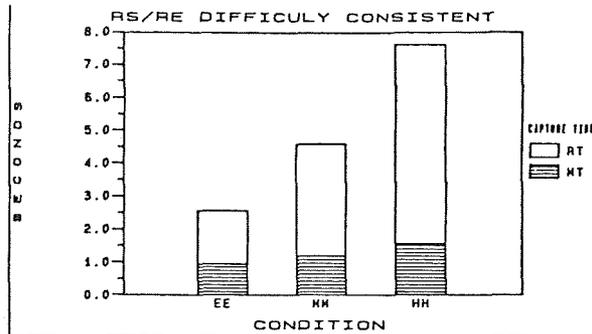


Figure 2b. RT-Before vs After for consistent conditions.

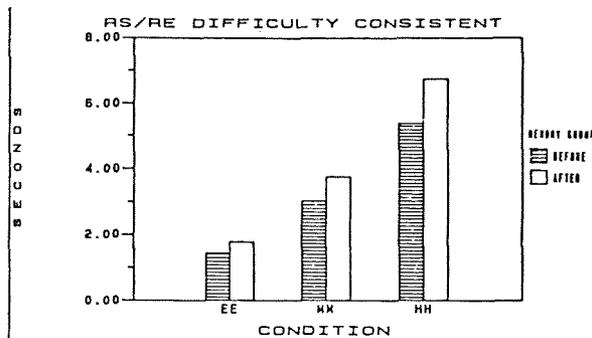
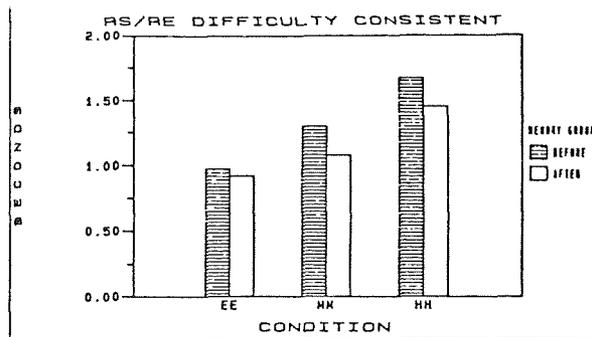


Figure 2c. MT-Before vs After for consistent conditions.



MTs increased as RE difficulty increased ($F(2,18) = 28.67, p < .001$).

Changing-consistent. As the RS/RE components increased in difficulty in the 'II' condition, and decreased in the 'DD' condition, RTs and MTs reflected the changing difficulty levels (Figure 3a). Before RTs were less than After RTs ($F(1,324) = 22.32, p < .001$), (Figure 3b), while their MTs were greater ($F(1,324) = 25.87, p < .001$), (Figure 3c).

Before groups. For this group, there was a significant interaction between conditions (II,DD) and segment for RT ($F(2,18) = 43.84, p < .001$). As RS difficulty increased across segments in the 'II' condition, and decreased in the 'DD' condition, the RTs increased or decreased respectively. MTs reflected the same interaction for the RE component ($F(2,18) = 52.16, p < .001$).

After groups. There was a significant interaction between conditions (II,DD) and segment ($F(2,18) = 62.76, p < .001$). As the RS difficulty increased across segments in the 'II' condition, and decreased in the 'DD' condition, RT increased or decreased respectively. Again, MT reflected the same interaction in the RE component ($F(2,18) = 29.67, p < .001$).

Changing-inconsistent. The difficulties of the RS and RE components for the 'ID', and 'DI' were varied in opposite directions. For the 'ID' condition, as the RS component increased in difficulty across segments within the condition, the RE component decreased in difficulty. The converse was true for the 'DI' condition. RT reflected the RS manipulations and the MT reflected the RE manipulations independently (Figure 4a). As in the previous two conditions, Before RTs were less than After RTs ($F(1,324) = 24.92, p < .001$), (Figure 4b), while their MTs were greater ($F(1,324) = 28.89, p < .001$), (Figure 4c).

Before group. There was a signifi-

Figure 3a. Capture time-RT vs MT for changing-consistent conditions.

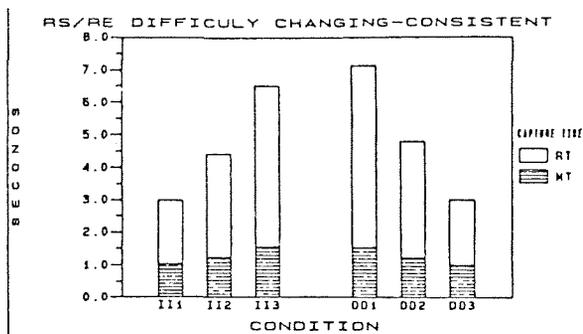


Figure 3b. RT-Before vs After for changing-consistent conditions.

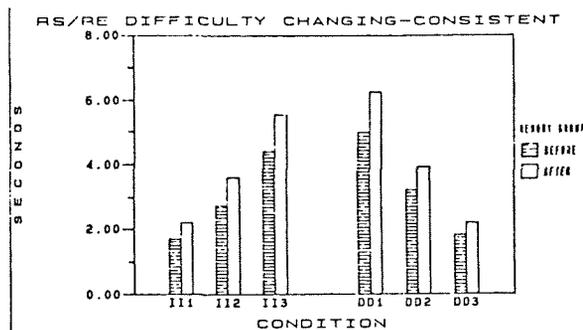


Figure 3c. MT-Before vs After for changing-consistent conditions.

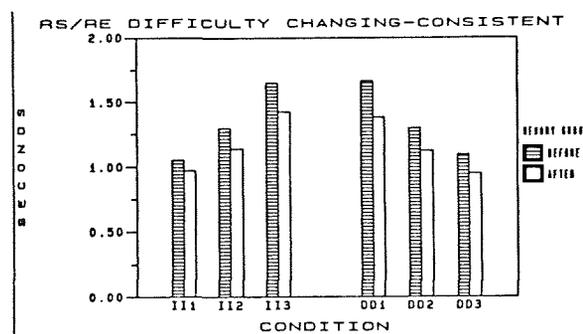
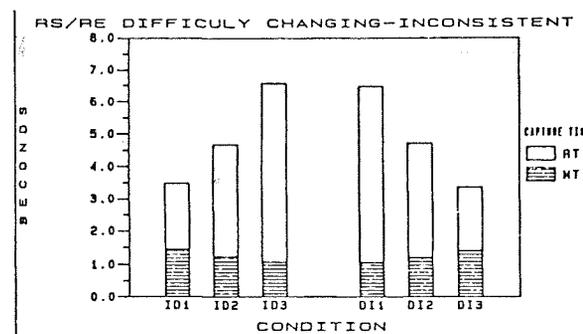


Figure 4a. Capture time-RT vs MT for changing-inconsistent conditions.



cant interaction between conditions (ID,DI), and segment. For the 'ID' condition, RTs increased as the RS component increased in difficulty, or decreased as the RS component decreased in difficulty ($F(2,18) = 36.60$, $p < .001$). Conversely, MTs decreased as the RE component decreased in difficulty in the 'ID' condition, and increased in the 'DI' condition ($F(2,18) = 37.98$, $p < .001$).

After group. The interaction between conditions and segment for RTs and MTs followed the same pattern as found in the Before group. RTs in the 'ID' and 'DI' conditions were inversely related ($F(2,18) = 104.74$, $p < .001$), as were the MTs in the same two conditions ($F(2,18) = 17.13$, $p < .001$).

Subjective Ratings

Relative importance of workload-related factors. There were large differences in the importance that subjects placed on the nine factors. Due to this variability in subject biases, there were no significant differences between memory groups in the relative importance each subject placed on the workload-related factors (Figure 5). These results follow widespread findings of variability in subjects biases, substantiating the importance of using weights to reduce between-subject variability in subjective evaluations of workload.

Weighted bipolar ratings Weighted bipolar ratings were weighted workload. Their means ranged from 19 to 49 for the Before group, and 8 to 50 for the After group. The workload involved in performing the 21 experimental conditions was evaluated at the end of each block of trials. These ratings were combined with the weights to calculate the weighted workload of the experimental tasks. This reduced between-subject variability by 32%. Once weighted workload was calculated, ANOVAs were conducted for the same three categories: (1) Consistent, (2) Changing-consistent, and (3) Changing-inconsistent. Separate ANOVAs were conducted for the Before and After groups. Weighted workload generally reflected the results obtained for the performance data.

Consistent (Figure 6). The Before group rated the RS/RE difficulty in the 'EE', 'MM', and 'HH' conditions as having significantly more workload than the After group did ($F(1,162) = 7.59$, $p < .01$).

Figure 4b. RT-Before vs After for changing-inconsistent conditions.

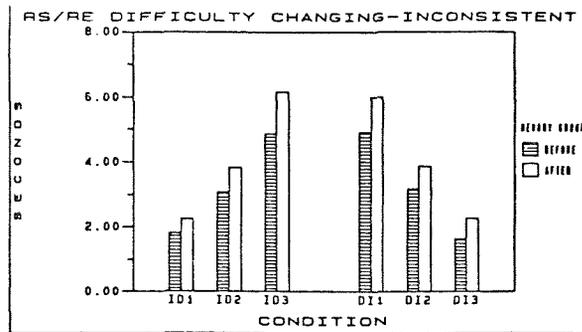


Figure 4c. MT-Before vs After for changing-inconsistent conditions.

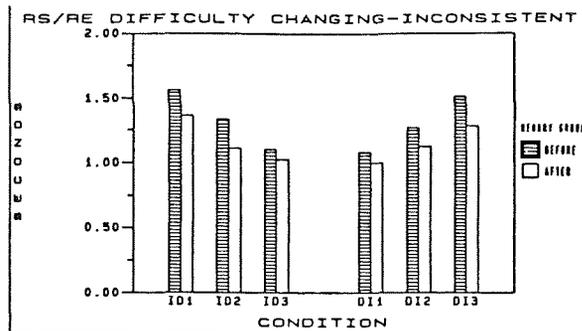
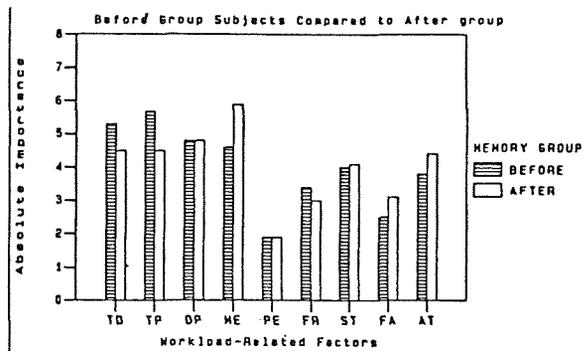


Figure 5. Relative importance of workload-related factors.



Before group. Workload increased from the 'EE' to 'MM' to 'HH' conditions as the RS/RE difficulty increased ($F(2,18) = 23.45, p < .001$). There was a small but significant effect between rated segments for the 'EE' condition ($F(2,18) = 4.97, p < .05$), but there were no significant effects between rated segments for the 'MM', and 'HH' conditions.

After group. Workload increased across conditions similarly to the increase in the Before group ($F(2,18) = 19.04, p < .001$). Within the 'EE' condition, there was a significant effect between rated segments ($F(2,18) = 4.05, p < .05$), but there were no significant effects between rated segments for the 'MM', and 'HH' conditions.

Changing-consistent (Figure 7). Subjects in the Before group rated the workload in the 'II' and 'DD' conditions higher the After group did, Figure 8. Weighted workload-Before but the differences were not vs After for changing-inconsistent significant conditions.

Before group. Workload ratings increased across rated segments within the 'II' condition ($F(2,18) = 4.09, p < .05$), and decreased across conditions within the 'DD' condition ($F(2,18) = 5.79, p < .05$).

After group. The results for the After group parallel those of the Before group. Ratings increased across rated segments within the 'II' condition ($F(2,18) = 6.01, p < .05$), and decreased across rated segments within the 'DD' condition ($F(2,18) = 3.07, p < .05$).

Changing-inconsistent(Figure 8). There was a significant difference in workload ratings between groups for the 'ID', and 'DI' conditions. Across segments, the Before groups ratings were greater ($F(1,162) = 4.25, p < .05$).

Before group. There were no significant effects, or interactions between conditions or segments for the 'ID', and 'DI' conditions. Workload ratings in the 'ID' condition did not reflect increased RS difficulty or decreased RE difficulty.

Figure 6. Weighted workload-Before vs After for consistent conditions.

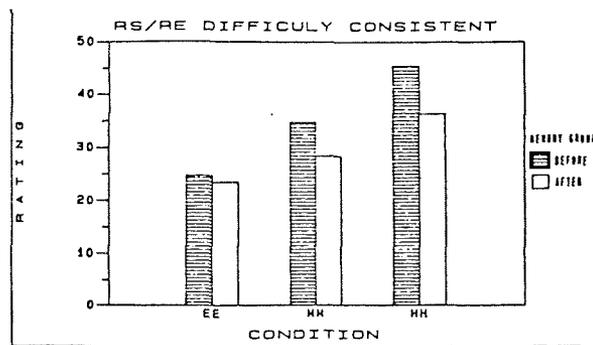


Figure 7. Weighted workload-Before vs After for changing-consistent conditions.

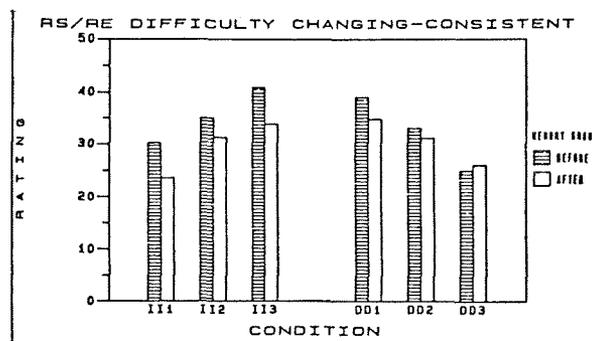
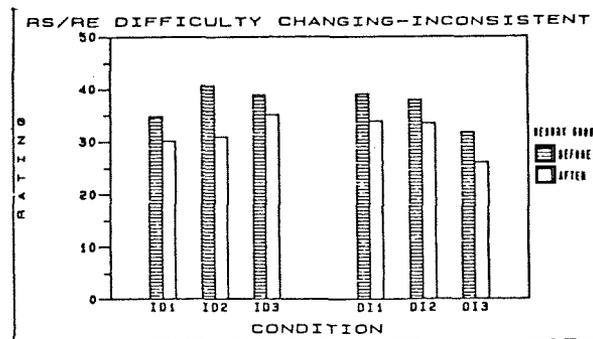


Figure 8. Weighted workload-Before vs After for changing-inconsistent conditions.



After group. Although marginally significant, the differences between rated segments in 'ID', and 'DI' conditions did not clearly reflect both RS and RE difficulty manipulations in an orderly way. For these conditions, workload ratings were more influenced by RS than RE.

Correlations among workload ratings and performance measures. Table 1 shows the correlations among the bipolar ratings, weighted workload, RT and MT, obtained with BMDP 6R. There were large variations in the correlations between raw bipolar ratings, and did not correlate very highly with RT and MT. With the exception of activity type, raw bipolar ratings highly correlated with weighted workload.

Table 1. Correlations among bipolar ratings, weighted workload, RT, and Mt.

	TD	TP	PF	ME	PE	FR	ST	FA	AT	OW	WW	RT	MT
Task Difficulty	-												
Time Pressure	.77	-											
Performance	.35	.35	-										
Mental Effort	.59	.56	.16	-									
Physical Effort	.73	.58	.28	.46	-								
Frustration	.70	.71	.46	.42	.64	-							
Stress	.71	.79	.15	.45	.55	.72	-						
Fatigue	.37	.44	.14	.13	.34	.44	.60	-					
Activity Type	.24	.13	.00	.23	.27	.12	.15	-.12	-				
Overall Workload	.86	.74	.23	.55	.69	.67	.72	.39	.25	-			
Weighted workload	.89	.83	.53	.62	.78	.81	.76	.52	.26	.79	-		
RT	.26	.11	.19	.13	.24	.14	-.05	-.03	.19	.18	.21	-	
MT	.42	.43	.40	.29	.29	.44	.35	.15	.08	.40	.43	.23	-

Discussion

The results of this experiment support the findings of previous experiments (Yeh, et al, 1985; Hart, et al, 1984; 1985) that subjects ratings are sensitive to task manipulations. Performance measures (RTs, and MTs) accurately and consistently reflected the difficulty manipulations in RS and RE components across the consistent conditions (EE,MM,HH). This supports earlier views that as cognitive loading increases as a function of increasing difficulty, performance measures increase. Performance measures also reflected the different difficulty levels in RS and RE when the difficulty within conditions was positively correlated, as in the 'II' and 'DD' conditions, or when the difficulty within conditions was negatively correlated, as in the 'ID' and 'DI' conditions. In all the conditions, RTs were driven by RS components, and MTs were driven by RE components. This is evident in the changing-inconsistent condition (ID, DI), where RTs varied with MTs the same way RS components varied with RE components. The fact that RTs were slower than MTs, suggests that the RS component, solving math equations, loaded cognitive processes more heavily than the RE component. These performance results hold true for the Before group, as well as the After group.

Subjective ratings also were sensitive to cognitive loading (Yeh et al, 1985; Hart et al, 1984), and reflect task manipulations. A major concern of this experiment was to look at the degree to which introspective subjective ratings were sensitive to specific variation in cognitive loading of segments

within a block of trials. Yeh demonstrated that subjects could integrate all the information in a block of trials, and could differentiate between different levels of cognitive loads with retrospective workload ratings. This experiment demonstrated that subjects are also sensitive to different cognitive loads within blocks of trials, although the degree to which retrospective workload ratings reflect manipulations in cognitive loading depends on the difficulty levels within conditions.

In the consistent (EE, MM, HH), and the changing-consistent condition ('II', 'DD'), the information about RS/RE difficulty levels was still available retrospectively, and workload ratings selectively reflected the difficulty of individual segments. In the consistent conditions, subjects rated segments of the same difficulty level as having the same workload. In the changing-consistent conditions, subjects rated segments of different difficulty levels as having significantly different workload. In this case, difficulty segments were rated as being more loading than medium difficulty segments, which were rated as being more loading than segments of easy difficulty. Knowing in advance did not appear to increase subjects sensitivity to task manipulations. Possibly subjects in the Before group gave higher workload ratings than subjects in the After group due to individual differences rather than increased sensitivity to the magnitude of difficulty manipulations, because the interactions between subject, group, experimental condition, and trial block segments were not significant. However, this difference may be due to a perceived additional memory task for the before group.

These results suggest workload ratings are a good indicator of the direction of RS/RE component difficulty manipulations rather than absolute magnitude, but only so long as the difficulty levels of the RS components and RE components were consistent and varied in the same direction. When this occurred, performing the RS/RE task components facilitated recall of the average difficulty of the task components for each segment of a different difficulty level. These findings are unlike dual-task results which reflect interference between tasks due to direct competition for limited resources. Since the output from the RS component serially fed into the RE component, and had to be completed prior to RE, the pairing of these processes did not lead to competition for common resources. Therefore, workload ratings reflecting the differences in difficulty between segments were reinforced.

In the changing-inconsistent condition (ID,DI), the difficulty levels of the RS and RE components were varied in the opposite directions. In this case, performing the RS/RE task components facilitated recall of the average difficulty of the task components across segments of different difficulty levels. It may be that more resources were allocated for integrating task components as in the changing consistent conditions, However, since the task components had opposing difficulty levels, recall of the average workload of the difficulty levels experienced across segments was facilitated. Consequently, workload ratings did not significantly reflect the direction of either the RS or RE component. This suggests that the workload ratings were not driven exclusively by the response selection component (which had a higher cognitive load than the response execution component), as RTs were, but by an integration of the two components. Although, in the 'ID', and 'DI' conditions for the After group, the workload ratings of the third segment reflected the difficulty level of the RS component, while the workload ratings of the first two segments reflected an integration of the two components. This appears to be a small recency effect, and suggests that the when integrating two task compon-

ents, RS may carry more weight, (i.e. the component that loads heavier on cognitive processes may weigh heavier in evaluating workload).

Conclusion

This study succeeded in determining some of the limits memory imposes on subjective ratings. Subjects appear to be sensitive to task component manipulations, and their ratings reflect the specific retrieval of information from long-term memory about the workload of particular segments, but only in certain conditions. Task components need to be stimulus/response compatible and well integrated for a human operator to accurately recall segments of a task that vary in difficulty, as all were in this study. If the task components vary in difficulty, human operators integrate them and recall the average workload of the difficulty levels. It appears that knowing in advance which segment should be rated may not additionally facilitate recall. Finally, the results from the changing-inconsistent condition indicate that the response selection component may load on cognitive processes more heavily, and consequently contribute more to workload ratings than the response execution component. Thus, the degree to which the response selection component drives workload ratings may be greater under some circumstances and not under others, and requires further research.

Acknowledgements

The authors wish to thank Mike Vidulich, Ron Miller, Jay Shively and Vern Battiste for their invaluable help and suggestions in running this experiment, and writing up the results. This work was performed under San Jose State Grant NCC-2-237.

References

- Arbak, C. J., Shew, R. L., and Simons, J.C. (1984) The use of reflective swat for workload assessment. Proceedings of the 1984 Human Factors Society Annual Meeting, 952-962.
- Bortolussi, M. R., Kantowitz, B. H., & Hart, S. G. (in press). Measuring pilot workload in a motion base trainer: a comparison of four techniques. Proceedings of the 1985 Human Factors Society Annual Meeting.
- Eggemeir, F. T., Melville, B. E., & Crabtree, M.S. (1984). The effect of intervening task performance on subjective workload ratings. Proceedings of the 1984 Human Factors Society Annual Meeting, 954-958.
- Hart, S. G., Sellers, J. J., & Guthart, G. (1984). The impact of response selection and response execution difficulty on the subjective experience of workload. Proceedings of the 1984 Human Factors Society Annual Meeting, 732-736.
- Hartzell, E. J., Gopher, D., Hart, S. G., Lee, E. & Dunbar, S. (1983). The Fittsberg law: the joint impact of memory load and movement difficulty. Proceedings of the 1983 Human Factors Society Annual Meeting.
- Notestine, J. C. (1984). Subjective workload assessment and effect of delayed

ratings in a probability monitoring task. Proceedings of the 1984 Human Factors Society Annual Meeting, 685-689.

Yeh, Y. (in press). The effect of varying task difficulty on subjective workload. Proceedings of the 1985 Human Factors Society Annual Meeting.